



Auditory model-based parameter estimation and selection of the most informative experimental conditions

Anna Dietze^{1,2,*} , Anna-Lena Reinsch^{1,2}, Jörg Encke^{1,2} , and Mathias Dietz^{1,2} 

¹ Department of Medical Physics and Acoustics, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

² Cluster of Excellence "Hearing4all", Oldenburg, Germany

Received 25 April 2023, Accepted 4 December 2023

Abstract – Identifying the causes underlying a person's hearing impairment is challenging. It requires linking the results of listening tests to possible pathologies of the highly non-linear auditory system. This process is further aggravated by restrictions in measurement time, especially in clinical settings. A central but difficult goal is thus, to maximize the diagnostic information that is collectable within a given time frame. This study demonstrates the practical applicability of the model-based experiment-steering procedure introduced in Herrmann and Dietz (2021, *Acta Acustica*, 5:51). The approach chooses the stimuli that are presented and estimates the model parameters best predicting the subject's performance using a maximum-likelihood method. The same binaural tone-in-noise detection task was conducted using two measurement procedures: A standard adaptive staircase procedure and the model-based selection procedure based on an existing model. The model-steered procedure reached the same accuracy of model parameter estimation in on average only 42% of the time that was required with the standard adaptive procedure. Difficulties regarding the choice of a reliable model and reasonable discretization steps of its parameters are discussed. Although the physiological causes of an individual's results cannot directly be inferred using this procedure, a characterization in terms of functional parameters is possible.

Keywords: Binaural hearing, Tone-in-noise detection, Computational audiology, Model-based experiment steering, Audiological diagnostics

1 Introduction

The aim of audiological diagnostics is to identify the causes of a person's hearing impairment. A broad range of measurement techniques covering all kinds of deficits in the auditory system is available (for a review see [1]). To achieve a good diagnosis, comprehensive test batteries including subjective and objective tests are usually carried out as a first step. While some measurements specifically test for a particular pathology, combinations of tests are often required to differentiate between causes. This linking of data to the underlying cause or pathology is then the second step of the diagnostic process, posing challenges for audiologists, ENT doctors, and researchers alike, for three main reasons. First, a variety of pathologies and their combinations can cause a similar outcome. Second, the realization that more data on a particular experiment or stimulus would have been required often comes subsequent to the data collection. At this point, obtaining more data is sometimes no longer practically possible and often inconvenient. But even if data would exist in abundance, a third

challenge remains: The auditory system consists of several highly non-linear stages intertwined with multiple efferent regulations. An experienced professional might be able to interpret the data and relate it to a unique pathology, but such diagnosis remains qualitative. A quantitative description of pathology-descriptive parameters with confidence ranges could provide information such as: The estimated loss of type I auditory fiber synapses range between 20% and 30%.

Computer models have been suggested as possible assistants in relating data to potential pathologies. Panda et al. [2] used a physiological model of the cochlea [3] to simulate data from a psychoacoustic test battery from hearing-impaired listeners. By varying one model parameter at a time, they created individualized computer models that enabled suggestions on underlying pathologies of their patients, although a combination of parameters would have yielded even better results in some cases. Model-based hearing diagnostics based on wideband tympanometry measurements was proposed by Sackmann et al. [4]. A finite element model of a human ear was used to simulate various pathologies like the stiffening of ligaments or joints to determine the most confident parameter set.

*Corresponding author: anna.dietze@uni-oldenburg.de

Comprehensive physiological models of the auditory system require a large number of parameters to be confined (e.g., [4, 5]). In addition, physiological redundancies and co-dependencies in the system are useful to stabilize auditory perception against small disturbances or minor impairments, but they also lead to ambiguities in confining model parameters (e.g., [6]). Functional models, on the other hand, require fewer, though more abstract, parameters, such as filter bandwidth, internal noise, or attenuation. For instance, Plomp [7] presented a quantitative model predicting speech understanding in noise that had only two parameters: attenuation and distortion. Confining these parameters does not lead to a description in terms of physiological characteristics. Nevertheless, such functional models can help with profiling hearing impaired persons and can predict the benefit to be expected from a hearing aid or hearing prosthesis. For instance, a prediction of common audiological functional parameters (CAFPAs) [8] from previously acquired audiological data using different machine learning algorithms has been presented in [9]. However, a large amount of data from different measurements is necessary.

The amount of experimental data required to confine the model parameters depends critically on two factors: Measurement accuracy (which depends on the square root of the number of trials) and the number of free model parameters (which causes a factorial effect on the number of parameter combinations). A single parameter can often be estimated from data obtained within a few minutes (e.g., [10]). Appraisal of three parameters, however, can already be expected to require several hours of data collection, at least in psychophysics (e.g. [11]). In many cases, it may be prudent to adjust the measurement, based on interim results. The approach of Sanchez Lopez et al. [12] for instance, can identify the most informative predictors in an auditory test battery, based on the preceding results. Instead of conducting all tests on each individual, only a subset of tests is sufficient for the characterization of listeners. These tests represent the nodes of a decision tree that lead to different diagnoses. Another way to confine the assessment of model parameters in a theoretically most time-efficient way is the maximum likelihood-based procedure running in parallel to the measurement and selecting those stimuli or tests that cause the best refinement in model parameters [11]. Theoretically, it can be used with any model and any portfolio of experiments. Nevertheless, the demands on the chosen model are high. It must provide good fits to all data without too many parameters. Otherwise, systematic deviations between model and data under any one experimental condition may cause the procedure to overemphasize this condition or to cause some other form of undesired behavior. Also, co-dependencies of the model parameters should be at a minimum.

The goal of the present study was to test the feasibility of model-based experiment steering for the prediction of model parameters. With this method, the experiment or the experimental conditions are varied such that prediction accuracy for diagnostically relevant model parameters is optimized. In contrast, standard adaptive methods choose

experimental conditions that optimize prediction accuracy in the dimension of the adaptive stimulus parameter (e.g., tone level in dB). A characterization of the model-based experiment steering method in real instead of simulated subjects as shown by Herrmann and Dietz [11] was performed. As we are working particularly on binaural aspects, a simple model of binaural hearing was used for the present proof of concept. The chosen model by Encke and Dietz can be fit to accurately simulate individual tone-in-noise detection sensitivity for stimuli that differ in interaural phase and noise correlation [13].

2 Methods

2.1 Model-based selection framework

The basis of the model-based experiment steering that is applied for this proof-of-concept study was presented by Herrmann and Dietz in [11]. It is a likelihood-based adaptive procedure that operates in the model-parameter space and provides estimations for model parameters that can then be used for diagnosis. In order to get the most diagnostic information, the stimulus is adaptively varied such that the accuracy of the model parameter estimation is maximized. The framework can be separated into two parts: A likelihood-based parameter estimation module, and an experiment steering module.

The parameter estimation module estimates those model parameters with which the model and the participant produce the most similar data. This parameter estimation module can also be used on data that was collected conventionally, i.e., without model-based experiment steering. For the analysis, all experimental data are compared with pre-calculated model predictions (stored in the so-called model table), that are based on a selected set of parameter combinations. The dimensionality of the model table equals the sum of N model and M stimulus parameters.

The comparison of experimental data and the model table yields a multi-dimensional likelihood space with high values representing a high likelihood of the data being generated by a specific combination of model parameters. Different features of this likelihood space can be of interest, depending on the specific research question or clinical task. We decided to get estimates for the most likely model parameter value and the accuracy of the prediction for each model parameter in isolation. Therefore, the $N+M$ -dimensional likelihood space is averaged over $N+M-1$ dimensions resulting in a compound likelihood distribution along the remaining parameter that was left out of the averaging. To derive the parameter estimation (mean, μ) and the accuracy of the estimation (standard deviation, σ) of this likelihood distribution we fit a Gaussian function to the distribution, with μ and σ as fit parameters. More precisely, for numerical convenience we fit a parabola function f of the form

$$f(x) = -\frac{(x - \mu)^2}{2\sigma^2} \quad (1)$$

to the log-likelihood values over the parameter values x . An offset parameter was not necessary, because the compound likelihood values were normalized by the maximal value, resulting a log-likelihood maximum of zero. The process is repeated N times, to fit the compound likelihood distribution of each of the N model parameters.

The second part of the framework is the model-based experiment steering (MoBES) module that runs in parallel to the data collection. It chooses the best experimental condition or the best stimulus to present to the subject next. With the MoBES module, the chosen stimulus is (based on the current model parameter estimates) expected to provide the most information for refining the model parameter estimates. The procedure chooses the stimulus condition that causes the largest reduction in σ . Within the framework, all model parameter values are discretized to simplify computation. Equation (1) can be applied to both continuous and discrete parameter values x . In the employed discrete version, x , μ , and σ can be expressed relative to the respective step size, i.e., in an arbitrary unit of “steps”. For each parameter the discretization step size should be chosen so that it corresponds to a small but measurable and diagnostically relevant difference. This guideline should also ensure that all parameter steps influence the simulated results by a similar amount, but of course for different experimental conditions. The scale on which discretization is performed (e.g., linear, logarithmic, or other) must be chosen such that the likelihood values over x are approximately normally distributed.

2.2 Experiment and auditory model

We chose a tone-in-noise detection experiment in which a tone (either interaurally in phase: S_0 or anti-phasic: S_π) has to be detected in noise. The interaural correlation of the noise (ρ) can vary from -1 to 1 , i.e., the noise is either anti-phasic ($N_{\rho=-1}$, referred to as N_π), interaurally fully correlated ($N_{\rho=1}$, referred to as N_0), or correlated to some extent in between these extreme conditions ($N_{-1<\rho<1}$). Detectability of the tone depends on its level and on the available interaural cues. The conditions without interaural cues (N_0S_0 and $N_\pi S_\pi$) are expected to be detected worst. Vice versa, detectability is expected to improve with increasing average interaural differences, being best for the conditions N_0S_0 and $N_\pi S_\pi$.

As noted in the Introduction, an accurate model is a crucial prerequisite for using the MoBES module. For this proof of concept, we opted for the analytic binaural processing model of Encke and Dietz [13]. It can predict correct rates of tone-in-noise detection for a variety of dichotic and diotic stimuli (such as the stimuli used in the experiment described above) with the three free parameters σ_{mon} , σ_{bin} , and $\hat{\rho}$. In the model the complex-valued correlation coefficient γ is calculated to quantify the amount of interaural phase difference (IPD) fluctuations, which is suggested to underlie binaural unmasking and is therefore used to estimate detectability. The model consists of a monaural and a binaural branch:

$$d' = \sqrt{d'^2_{\text{mon}} + d'^2_{\text{bin}}}. \quad (2)$$

The monaural branch is sensitive to differences in energy between the reference and the target signal:

$$d'_{\text{mon}} = \frac{\text{SNR}_{\text{eff}}}{\sigma_{\text{mon}}}. \quad (3)$$

The sensitivity is inversely proportional to model parameter σ_{mon} . The binaural branch is based on the difference between the Fisher’s z -transformed complex correlation coefficients of a reference signal and a target signal:

$$d'_{\text{bin}} = \frac{|z[\hat{\rho}\gamma_r] - z[\hat{\rho}\gamma_t]|}{\sigma_{\text{bin}}}. \quad (4)$$

Since this transformation would result in infinite sensitivity to divergence of a fully coherent signal, which is not observed in the auditory system, the parameter $\hat{\rho}$ ($0 < \hat{\rho} < 1$) was introduced before z -transformation, thus limiting maximum sensitivity. As in the monaural branch, a model parameter σ_{bin} is used which is inversely proportional to binaural sensitivity, i.e., to the Euclidian distance between the z -transformed complex correlation coefficient of target and reference. The chosen experiment and model serve as one example use case of the MoBES procedure. Therefore, the model is only summarized here. Details can be found in [13].

Using this model as presented in [13], predicted detection thresholds are the same for detecting anti-phasic tones within diotic noise (N_0S_π) and for detecting in-phasic tones in anti-phasic noise ($N_\pi S_0$). This is not the case in behavioural data as shown for instance in [14]. We therefore modified the original model by introducing a fourth parameter into the model. It represents the decrease in $\hat{\rho}$ with increasing IPD, i.e., with increasing the argument of the complex correlation coefficient.

Introducing this additional parameter with a fading between the two most extreme correlation conditions of $+1$ (IPD = 0) and -1 (IPD = π) causes a slightly altered model architecture. The parameter that is limiting maximum sensitivity ($\hat{\rho}$ in Eq. (4)) is replaced by a term containing the parameter l_{max} limiting maximum sensitivity at IPD = 0 and the new parameter Δl_{max} that is representing the difference in sensitivity between the noise correlations of $+1$ and -1 :

$$\hat{\rho} = 1 - 2^{l_{\text{max}} + \frac{(1-\rho)\Delta l_{\text{max}}}{2}}. \quad (5)$$

Since the complex correlation coefficient has no imaginary part in our experiment, the real part of the Pearson correlation coefficient ($\rho = \Re\{\gamma\}$) of the reference signal and of the target signal is used instead.

Model predictions are shown in Figure 1. In each panel, one model parameter was varied, while the other three parameters were set to a fixed value in the center of their respective range. As described above, each model parameter introduces changes to specific stimulus conditions, whereas others are not affected.

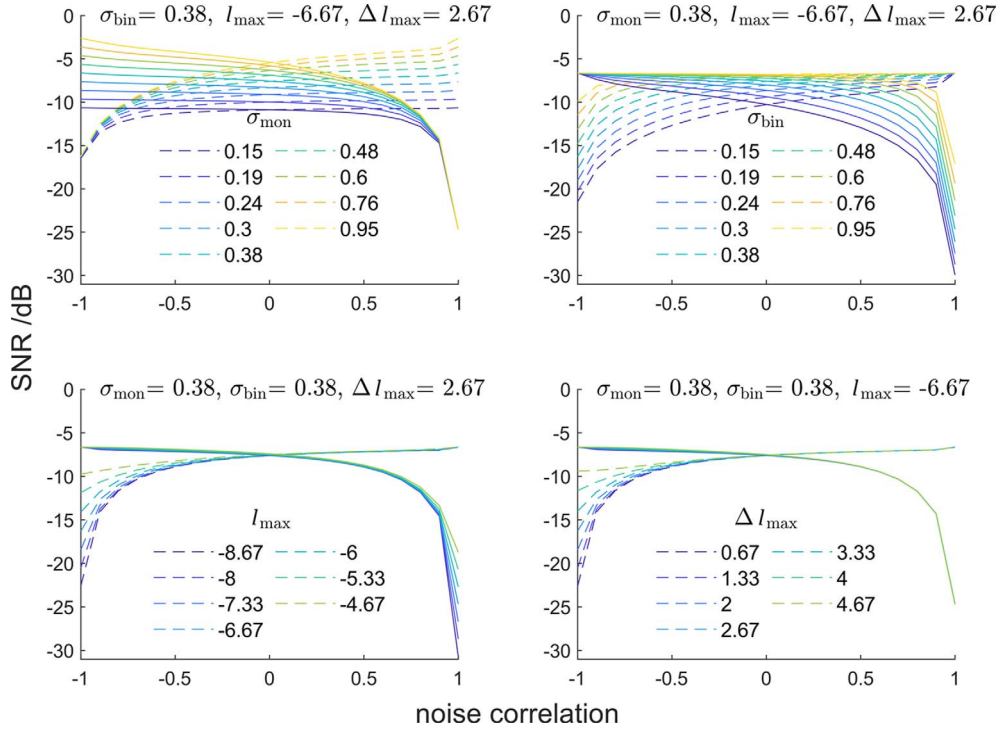


Figure 1. Model predictions (i.e., SNR corresponding to 79.4% correct) for different noise correlations (dashed lines: $N_\rho S_0$, solid lines: $N_\rho S_\pi$). In each panel, one model parameter was varied (color coding), while the other three parameters were set to a fixed value in the center of their respective range (shown at the top of each panel).

2.3 Measurements

Five young participants (age: 20–26 years; 3 female, 2 male) conducted the experiments with informed consent (approved by the ethics committee of the University of Oldenburg). The listeners received monetary compensation for the time spent on the experiments. Self-reported normal hearing was verified by clinical pure-tone audiometry (AT900, Auritec, Hamburg, Germany). None of the listeners had hearing thresholds exceeding 20 dB HL and there was no more than 10 dB difference in hearing threshold between the two ears at any octave frequency between 125 Hz and 10 kHz. The experiments were preceded by a training phase to familiarize the participants with the task. Two listeners had prior experience in binaural listening tasks (S1 and S5), the remaining three had no previous training in binaural hearing experiments.

2.3.1 Tasks and stimuli

The study consisted of two parts. All subjects participated in the same tone-in-noise detection task using (1) an adaptive staircase procedure and (2) the MoBES procedure. A four interval, two alternatives forced-choice experiment was conducted. Three intervals contained only the noise with a bandwidth of 100 Hz (Gaussian white noise with rectangular power-spectral density), arithmetically centered around 250 Hz. The second or third interval additionally contained a pure tone. This pure tone of 250 Hz was either interaurally in phase (S_0), or anti-phasic (S_π). The noise's

interaural correlation ρ ranged from anti-correlated to fully correlated (-1 , -0.75 , -0.5 , 0 , 0.5 , 0.75 , 1). The stimuli were chosen to be comparable to those used in Robinson and Jeffress [15]. The duration of the stimulus intervals was 0.6 s, each separated by 0.2 s silence intervals. A cosine rise-and-fall window of 20 ms was applied to the noise and to the pure tone separately. The tone started when the noise was at full amplitude. The level of the noise was fixed at 67 dB SPL, whereas the tone level was varied adaptively during both experiments, as described below.

The listeners sat in a sound-attenuating booth on a comfortable chair in front of a computer screen and a computer keyboard. The signals were transmitted to an external audio interface (ADI-2 DAC FS, RME, Heimhausen, Germany) and presented using circumaural headphones (HD650, Sennheiser, Wedemark, Germany). Four rectangles lit up on the screen in succession during the four intervals in order to visually support the temporal sequence. The participants' task was to decide whether the second or the third interval differed from the first and last "cueing" intervals. Responses could only be given after the fourth interval and were entered by pressing the number "2" or "3" on the keyboard. The button press was followed by visual feedback on the screen indicating whether the choice was correct. After a delay of 250 ms, the next trial was presented.

2.3.2 Adaptive staircase procedure

The first portion of the experiments was a standard adaptive staircase procedure varying the tone level

following a 1-up 3-down rule converging to 79.4% correct responses [16]. The initial step size of 6 dB was halved to 3 dB after the second and again to 1.5 dB after the fourth reversal. The 1.5 dB step size was used for eight reversals. Complete runs under the 14 unique stimulus conditions (seven noise correlations, each with two different tone IPDs) were presented in random order, one complete run after the other. Each condition was repeated five times. Whenever feasible, a complete set comprising all these 14 conditions was measured on the same day. These five sets will be referred to as the five “measurement sets”.

After completion of data collection, the likelihood-based parameter estimation module was applied to assess the most likely model parameters underlying these results. For visualization of the measured data, and for a comparison with the model predictions of the parameter estimation module, detection thresholds corresponding to 79.4% correct responses were computed from the average of the last eight reversals of the adaptive tracks.

2.3.3 Model-steered procedure

In the second part of the experiment, the measurement was conducted with the MoBES module introduced above. The range and discretization steps of the model parameters needed to be confined prior to the measurement phase.

Depending on how the parameters influenced the model outcome, the relation between the possible values was chosen differently. For σ_{mon} and σ_{bin} , factorial steps of $\sqrt[3]{2}$ ranging from 0.15 to 0.96 were chosen. Ranges from $-26/3$ to $-14/3$ for l_{max} and $2/3$ to $14/3$ for Δl_{max} were chosen with linear steps of $2/3$. The discretization was chosen for each parameter such that changes by one step led to approximately the same change in the SNR estimates. The effects of changes in each of the model parameters are shown in Figure 1. Changing σ_{mon} by one step always leads to changes in the estimated signal-to-noise ratio (SNR) of about 1 dB. Similar changes are observed for σ_{bin} but in other stimulus conditions. Increasing or decreasing parameters l_{max} and Δl_{max} by one step always leads to a change of about 2 dB but influences fewer stimulus conditions. Several piloting trials were necessary to ensure that the individual parameters of each subject were covered by the range of tested model parameters.

The model was run for all combinations of possible model parameters (model instances), and all combinations of possible stimulus parameters (stimulus conditions). The model table was pre-calculated overnight on a regular i5 laptop. The combination of the 2×7 stimulus conditions and the $9 \times 9 \times 7 \times 7$ model instances led to a total of 55,566 model calls for each of the 131 simulated stimulus levels to generate the psychometric functions. Instead of working with the original psychometrics functions (detection thresholds for different SNRs) for each stimulus, the amount of data in the model table is reduced by fitting a logistic function to the psychometric functions generated by the model for each combination of model instance and stimulus condition. The thresholds and slopes were obtained by a likelihood-based comparison of the psychometric functions generated by

the model and logistic functions with a wide range of possible thresholds and slopes. The thresholds and slopes of the logistic functions with the best fit were saved in the model table. During the MoBES procedure and when using the parameter estimation module for pre-recorded data, only the model outcome stored in this model table (thresholds and slopes) was available for the likelihood-fitting.

With human subjects, unlike artificial subjects, switching between perceptually differing stimulus conditions across single trials leads to less reliable responses and poorer immediate performance (e.g., [17]). To circumvent this, two additions were made to the original procedure: First, the measurement phase was split into several measurement blocks, each with a fixed number of trials of the same stimulus condition (but varying level corresponding to the point of maximal expected information). For this study, 28 blocks, each containing 30 trials of the same condition, were completed by the subjects. After each block of 30 trials, the MoBES module computed the next stimulus condition to be presented. Second, the first two trials of each block were carried out merely to permit familiarity with the new stimulus condition but were neither saved nor used for the steering procedure. With this, a total of 840 trials ($28 \text{ blocks} \times 30 \text{ trials}$) were presented, of which 784 trials ($28 \text{ blocks} \times 28 \text{ trials}$) were stored.

The first four blocks were measured under predefined conditions before the likelihood-based measurement steering algorithm started. This was to initialize the model with a good starting point for the selection of the subsequent stimulus conditions. The conditions chosen for these initial blocks were: one purely diotic condition ($N_0 S_0$), the two extreme dichotic conditions ($N_0 S_\pi$ and $N_\pi S_0$), plus one intermediate condition ($N_{\rho=0.75} S_\pi$). The choice of suitable initialization blocks also required knowledge acquired during the piloting of the study.

With the MoBES module, the accuracy of the model parameter estimation can be tracked and then used to terminate the experiments. With such a termination criterion, the measurement ends when the desired confidence range is reached for all model parameters. For the present “proof-of-concept” study, no termination criterion was set. Instead, a fixed number of 784 trials were conducted. This number was chosen to allow for comparisons between the two procedures, as the number of trials in one measurement set in the adaptive procedure was approximately 750 (depending on measurement set and subject).

3 Results

3.1 Adaptive staircase procedure

The tone-in-noise detection thresholds corresponding to 79.4% correct obtained with the five measurement sets of the adaptive staircase procedure are shown in Figure 2 where each panel shows data for one of the five subjects. Using the parameter estimation module, model parameters corresponding best to the subjects’ data were obtained. The resulting model predictions for the $N_\rho S_0$ and $N_\rho S_\pi$ conditions are displayed as dashed and solid lines in

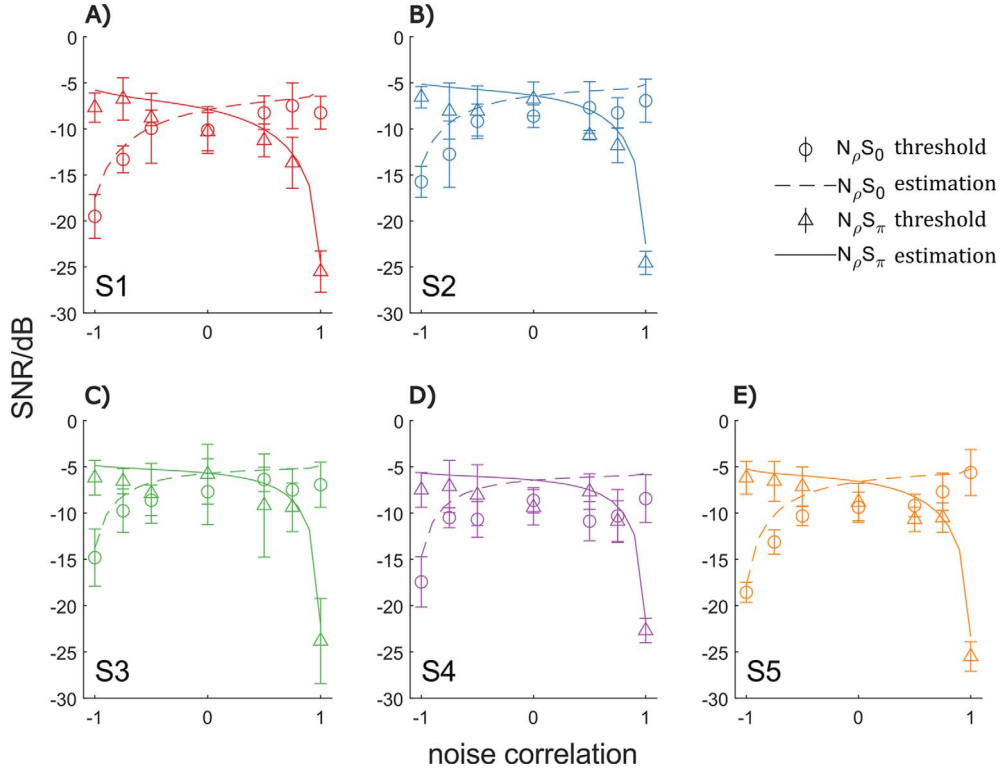


Figure 2. Tone-in-noise detection thresholds of the five subjects obtained with the adaptive staircase procedure. The triangles represent median thresholds for stimuli with anti-phasic tones ($N_\rho S_\pi$), the circles for tones that were inter-aurally in phase ($N_\rho S_0$). The inter-quartile range of the five trials of each condition is represented as error bars. The dashed lines (anti-phasic tones, $N_\rho S_\pi$) and solid lines (in-phase tones, $N_\rho S_0$) represent the SNR thresholds predicted by the model with the parameters estimated by the parameter estimation module.

the same figure and show the modelled SNR for 79.4% correct.

As expected, the thresholds for the conditions without binaural cues ($N_0 S_0$, the right-most circle and $N_\pi S_\pi$, the left-most triangle) are the highest. Thresholds improved with increasing average IPD difference between masker and target, until the lowest thresholds were obtained for $N_\pi S_0$ (leftmost circle) and $N_0 S_\pi$ (rightmost triangle). Within the latter condition, all subjects reached the lowest of their thresholds.

The model predictions of the parameter estimation module captures the behavior of all subjects, with only small deviations for single stimulus conditions (see Fig. 2). The performance is slightly underestimated by the model predictions in the conditions with the worst behavioral thresholds. The coefficient of determination R^2 ranged between 0.64 for subject S4 and 0.85 for subject S5 and was averaged 0.80. The SNR thresholds obtained with the adaptive procedure for the conditions $N_0 S_0$ and $N_0 S_\pi$ are shown in Figures 3A and 3D. Estimates for the four model parameters based on the five measurement sets individually (“1”, “2”, “3”, “4”, “5”, and the median of these results: “ μ ”) and all data analyzed together (“all”) are shown in Figures 3B, 3C, 3E, and 3F. It becomes obvious that the detection thresholds and model parameter estimates differ between the five measurement sets. The SNR of the $N_0 S_0$

condition drops over time, which is reflected in model parameter σ_{mon} and to some extent also in σ_{bin} . The variability in the model parameters l_{max} and Δl_{max} seems not to follow any systematic trend. The SNR estimated from all five measurement sets together differs for many conditions from the median SNR of the five adaptive measurement sets analyzed individually. This can be seen for instance in Figure 3A when comparing the circles to the dots. The difference ranges up to 5 standard deviations of the adaptive measurement sets. The above-mentioned difference is not reflected in the model parameters estimated from the five measurement sets together (circles) and the median of the five individual measurement sets (dots). For the model parameters the maximal difference is 0.6 standard deviations.

3.2 Model-steered procedure

When using the MoBES module, model parameters were estimated for every trial based on the compound likelihood for each model parameter. Figure 4 shows the development of the compound likelihood (mean over the other three parameters and all stimulus parameters after setting the maximum of each trial to zero) for each of the four parameters over trials for subject S4 in the upper four panels. Over the course of the trials, the likelihood

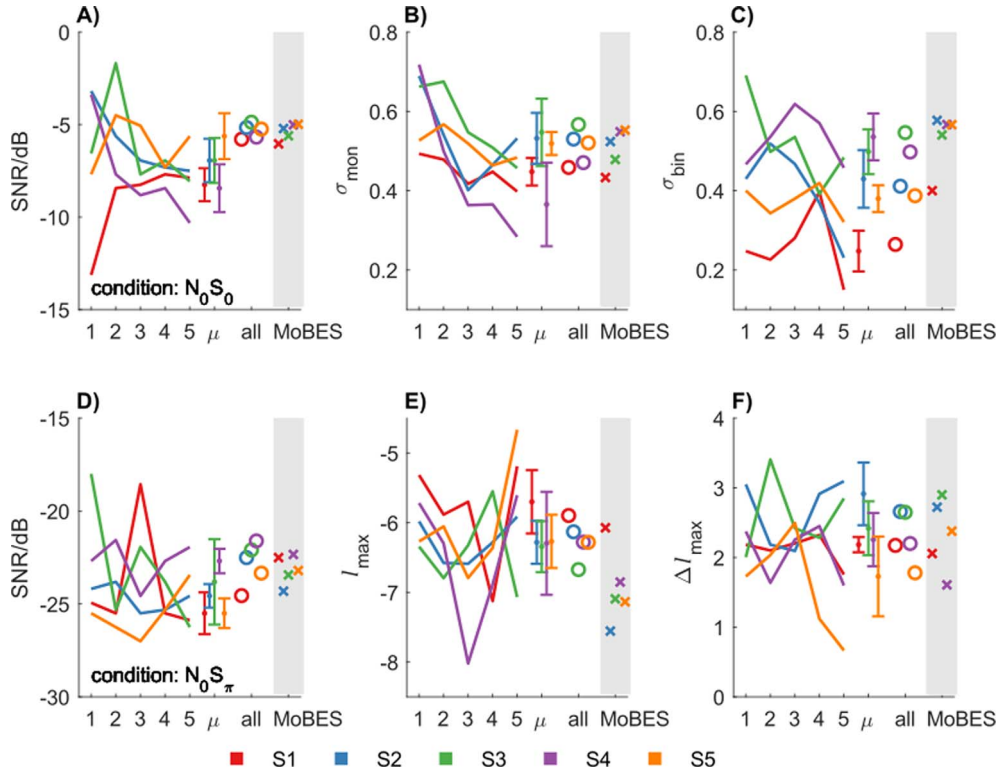


Figure 3. Detection thresholds (SNR, panels A, D) obtained with the adaptive procedure and model parameter estimates (panels B, C, E, F) determined with the parameter estimation module for the adaptive experiment. The lines represent the data for the five individual measurement sets. Their medians and inter-quartile ranges are shown with the dot and the error bars. The circles show the model parameter estimates for running the estimation module for the data of all five sets together. The model parameters for the data obtained with the MoBES procedure are indicated by the crosses above the gray shading.

distribution reduced in width. The bottom panel shows the stimuli chosen by the procedure.

Lower values of σ_{mon} correspond to lower thresholds in the diotic (or monaural) conditions. Lower values of σ_{bin} correspond to lower thresholds in those conditions with interaural differences. As described in the Methods section, the parameter l_{max} mainly affects the thresholds for $N_{\pi}S_0$ and N_0S_{π} , whereas Δl_{max} influences the difference between $N_{\pi}S_0$ and N_0S_{π} . This can be observed in Figure 4: The first stimulus condition in the experiment (N_0S_{π}) did not deliver information on the monaural threshold. For this reason, the estimation of σ_{mon} only starts refining with the second block (N_0S_0). Similarly, parameter Δl_{max} (the difference between N_0S_{π} and $N_{\pi}S_0$) can only be estimated starting with the first trials of $N_{\pi}S_0$ in block number three. After the four initialization blocks were presented, starting with trial number 113 (at the dashed black line), the experiment steering module selected different stimulus conditions, emphasizing N_0S_{π} , $N_{\rho=0.75}S_{\pi}$, and, to a lesser degree, $N_{\pi}S_0$, and $N_{\pi}S_{\pi}$. Noise correlation values between those were only rarely chosen (once in S1 and S4, twice in S2, and never in S3 and S5). Comparable patterns and similar model parameter estimates were also found for the other subjects (see Supplementary Figures 1–4). The estimates for the four model parameters based on data from the MoBES

procedure are shown in the grey shading of Figures 3B–3F and in Figure 5.

The mean of the confidence ranges (variance of the parabola fit, σ in Eq. (1)), which can be qualitatively estimated for subject S4 from the width of the likelihood surfaces for the four parameters in Figure 4, is shown for all subjects in Figure 6. As a global trend, the confidence ranges decreased with the number of trials. For instance, the mean confidence range over the four parameters for subject S4 decreased from 2.14 steps at the start of the model-steering to 0.36 steps after the last trial. Comparable decreases were also found for the other subjects. After the final trial, the procedure reached a mean accuracy between 0.32 steps and 0.36 steps for the different subjects (mean: 0.35 steps).

3.3 Comparison of the two procedures

The detection thresholds (SNR) and the model parameter estimates with the data acquired using the adaptive and the MoBES procedure are shown in Figure 3 (panels A, D and panels B, C, E, F, respectively). The SNR thresholds calculated from all five measurement sets of the adaptive procedure (circles) and those obtained by the MoBES procedure (crosses) are very similar. The differences in

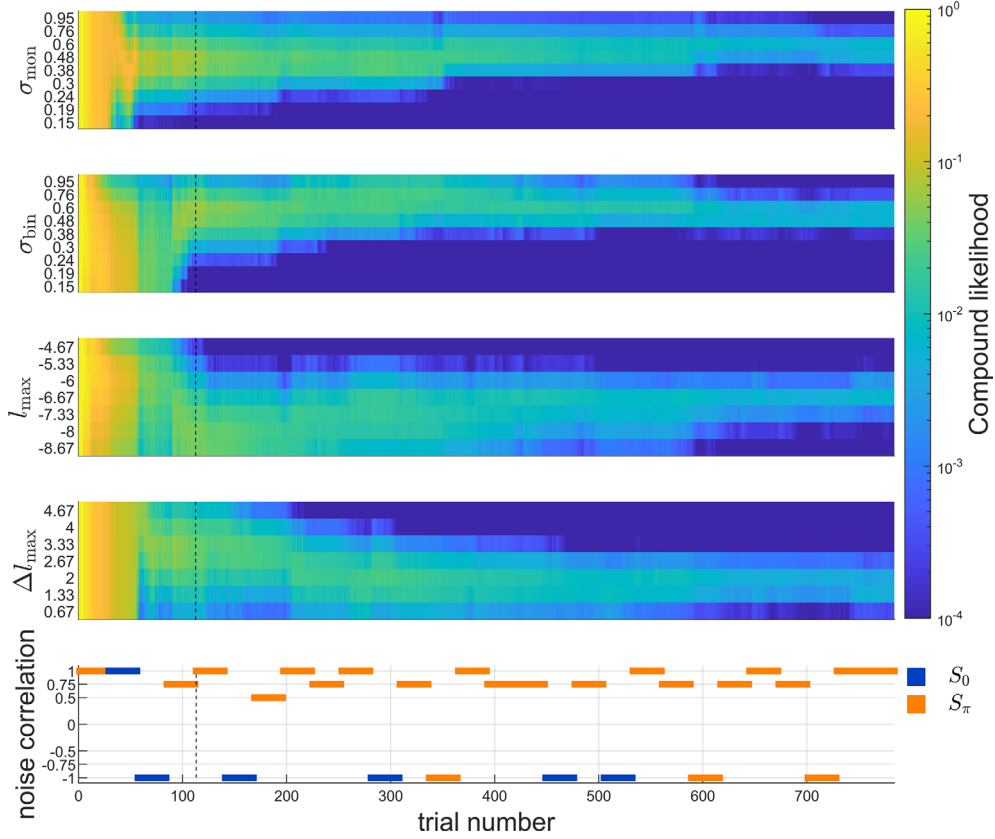


Figure 4. Compound likelihood (mean over the other three parameters and all stimulus parameters after setting the maximum of each trial to zero) for each of the four model parameters in the upper four panels for subject S4. The stimuli chosen by the procedure across trials are shown in the bottom panel. The dashed black line indicates the end of the initialization blocks and the start of the model steering with trial number 113.

SNR ranged between 0 and 2.0 times the standard deviation of the five adaptive measurement sets in all subjects and conditions with two exceptions: The estimates differed in subject S5 for the condition $N_\pi S_0$ by 3.0 and for subject S2 for the condition $N_0 S_\pi$ by 2.5 standard deviations.

As shown in Figure 3, the difference between the model parameters estimated from data obtained with the adaptive procedure (circles) and the MoBES procedures (crosses) ranges in all but two cases between 0.1 and 1.6 times the standard deviation of the parameters across the five adaptive measurement sets. Only in subject S2 (blue) the estimates from the two procedures differ by 4.5 standard deviations for parameter $l_{\max}$ and in subject S5 (orange) by 4.4 standard deviations for parameter σ_{bin} . For a better comparability and to detect possible biases of the two procedures, estimations of the four model parameters for the two procedures are shown in Figure 5. It becomes evident that the difference between the estimations by the two procedures depends on the specific model parameter and subject. However, a bias towards lower estimations for one procedure might be present for the parameters σ_{bin} and $l_{\max}$. In general, parameter estimations by the two procedures are comparable and the model parameter estimations do not differ substantially between the subjects.

Figure 6 shows the mean confidence ranges across the four model parameters as a function of trials for the

adaptive procedure and the MoBES procedure. With the latter, only 874 trials were recorded. For the adaptive procedure, confidence ranges are only shown after each full measurement set of 14 conditions (651–786 trials). These measurement sets differed slightly in the number of trials as the number of trials needed for eight reversals at the final step size differed between the subjects and measurement sets. In general, a decrease of confidence ranges over trials was observed, with a steeper decrease in the model-steered data. For instance, for subject S1, the mean confidence range after the first measurement set of the adaptive procedure (693 trials) was 0.82 steps. Using the MoBES module the same or a smaller value was reached after 302 trials. The same confidence range was achieved more than twice as fast with the MoBES module. To reach the same confidence range using the adaptive procedure, 1.9–3.7 times more trials were necessary than with the model-steering procedure.

4 Discussion

This study sought to test the feasibility of model-based experiment steering in human subjects, after a preceding study by Herrmann and Dietz [11] had concluded that there would be a theoretical advantage of the proposed procedure

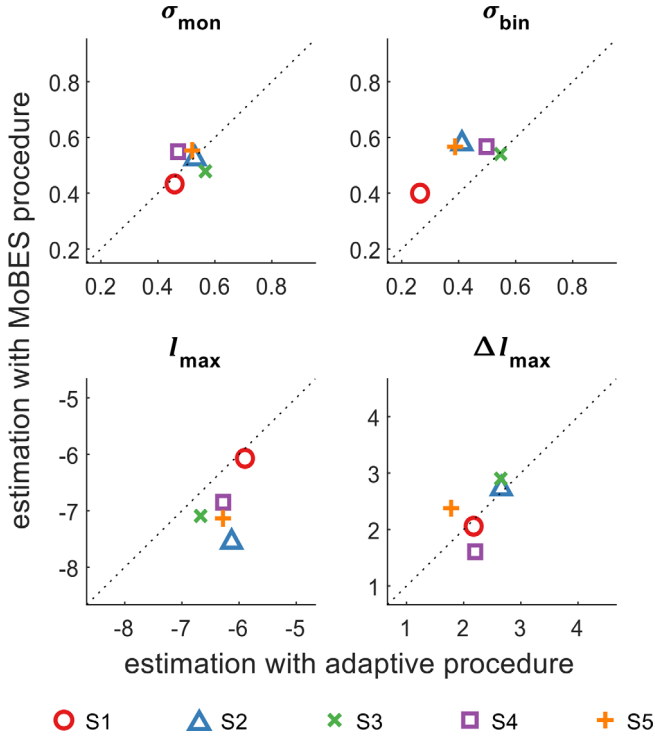


Figure 5. Estimations for the four model parameters based on all data from the standard adaptive procedure and the MoBES procedure for subjects S1 to S5. The dashed line indicates equal estimations based on the two procedures. Color and marker shape vary for the individual subjects.

over sequential measure-and-fit approaches. In the current study, the model-steered procedure was tested on young normal-hearing subjects, while the previous study only tested an artificial “in-silico patient”. This attempt was successful for two reasons. First, the estimated model parameters were sufficiently close to those obtained from the results of the standard adaptive procedure. Second, the same accuracy in model parameter estimates was obtained in 27–60% of the time required by the standard adaptive method. The proposed measurement procedure can assist in linking data to the underlying pathology, or to a parametric description of the individuals’ hearing abilities. The procedure will steer towards those measurements that can disentangle different causes of the observed behavior, even in the complex auditory processing chain. As a prerequisite for this becoming reality in clinical settings, models with high diagnostic resolution need to be developed. In the current study, an existing simple model of binaural processing was used, but slightly adapted as a first attempt to characterize a subject in the most time-efficient way. Even though the diagnostic value of the model parameters is not clear in this study, it served as proof of concept.

The duration of measurements is limited in clinical settings. However, keeping measurement times as short as possible is also of importance for another reason: With longer measurement times, unaccounted factors could influence the data. Fatigue, attention, motivation or effects specific to single measurement days may potentially

confound the parameter evaluations. Using the model-steering procedure, mean confidence ranges of 0.35 steps were reached for the four model parameters after less than 1.5 h. To put it another way, the model-steered procedure reached the same accuracy of model parameter estimation on average in 42% of the time required by the standard adaptive procedure. Attempts to shorten measurement times in clinical settings have been presented before (e.g., fast audiometric testing presented in [18]). In contrast to previous studies, the present study aimed for a procedure to reduce measurement times that is not restricted to a specific experiment.

One of the main concerns remains the choice of an accurate model with diagnostic value. The approach with an auditory processing model requires the faithful simulation of the whole chain from stimulus presentation, through internal processing, to the subject’s response, or to other measured data. We were able to perform a proof-of-concept but could only characterize those aspects that are relevant for tone-in-noise detection sensitivity at one frequency and only for normal-hearing subjects. The four model parameters cannot be directly related to hearing difficulties. The parameters σ_{mon} and σ_{bin} describe general monaural and binaural abilities of the participant. Importantly, these two parameters are both influenced by disturbances at various levels of the auditory system. Disturbances can range from conductive hearing loss and hair cell loss to cognitive factors such as attentional deficits. Parameter l_{max} is related to the best performance achieved by binaural hearing. Therefore, it is not fully independent of σ_{bin} . The physiological basis of Δl_{max} (difference in the firing rates to the conditions $N_{\pi}S_0$ and N_0S_{π}) is described in [19] but can also not be based on one process alone. This overlap in causes and effects is common in functional models. However, to diagnose the causes of hearing difficulties, other models are needed.

In order to use the approach with hearing impaired subjects, additional model parameters must be allowed to vary. For example, the parameter “effective bandwidth of the auditory periphery” is fixed in our model. Therefore, it cannot serve as a realistic model for patients with outer hair cell damage. Of course, this bandwidth could be an additional parameter to fit, as already demonstrated in [11], and most other specific extensions are also expected to be compatible with the approach. The problem is the number of parameters, especially as many of the parameters may differ from frequency to frequency. At the same time other parameters, such as the endocochlear potential are inherently frequency-independent, but influence hearing differently across frequency [2], further complicating a comprehensive parameterization. Abstract models that even avoid a simulation of auditory processing may be more realistic candidates for model-steered profiling. Abstract models can be employed if, instead of a detailed diagnosis, the focus of interest is rather on the consequences of altered auditory processing in real-world listening scenarios. Ideally, each model parameter should directly relate to a practical outcome, e.g., it can be a hearing-aid fitting parameter (similar to the model used by Plomp [7]).

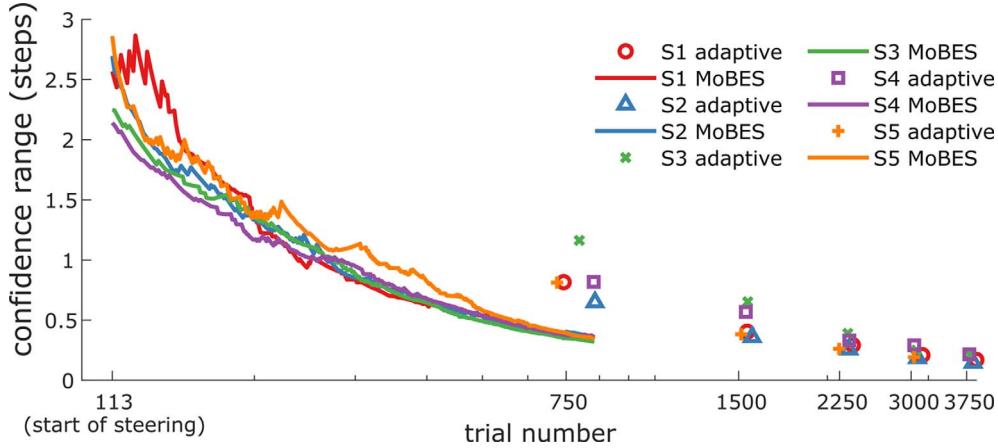


Figure 6. Mean of confidence ranges (in steps) averaged across the four model parameters over trials for subjects S1 to S5. The estimates for the model-steered procedure are depicted with lines, for the adaptive procedure with symbols. Color and marker shape vary for the individual subjects.

Having decided on a particular model, choosing meaningful ranges and discretization for the model parameters remains a critical point. In the best case, each step leads to similarly large changes in model predictions as shown in Figure 1. Matching the effect size of parameter steps is also important in the light of co-dependencies between model parameters. Preferably, changes by one discrete step in one parameter should not force another co-dependent parameter to change by more than one step. It is also important that estimated parameters do not reach the boundary of the parameter range of the previously stored model table. To fit the data best, the apex of the parabola that is used to obtain parameter estimation and confidence range, would possibly be outside the boundaries. The steepness would be very small, resulting in confidence ranges spanning the entire possible range of parameters. Such corrupted confidence ranges lead to the choice of non-optimal next stimulus conditions. An additional advantage of matching the effect size of parameter steps is that it allows the steering procedure to minimize the unweighted sum of confidence ranges, as measured in numbers of steps. The procedure is then expected to provide similar accuracy for all parameters without being biased towards minimizing the confidence ranges of some model parameters more than others. Extensive piloting with adjustments to the ranges and step sizes of the parameters preceded data collection. The need for such time-consuming preparation makes the method feasible only when the subsequent measurements benefit substantially from it. This is the case, for example, when many participants are to be measured (i.e., the extensive piloting time is outweighed by considerable savings in measurement time) or when the measurement time with these participants is restricted very much (i.e., the method allows for a better use of the limited time). Both are often the case in clinically-oriented patient studies.

Independent of the exact experiment or the population that is measured, at least two sources of variability can

influence the data in behavioral measurements. First, the variability of responses over time, which can be influenced by training, fatigue, attention, motivation, and other factors. Second, the pathology-induced changes to the system that we aim to quantify in terms of model parameter estimates. A training effect was observed for the standard adaptive measurements. The SNR thresholds were significantly lower in the fifth measurement set compared to the first measurement set in all but one participant. This training effect can be seen for some participants in the two conditions shown in Figures 3A and 3D and might be present within the results of the model-steered experiment, too. Importantly, the MoBES procedure does not operate in the dimension of threshold values as those procedures reviewed in Leek [20], but in the dimension of model parameters, trying to minimize the confidence ranges of the model parameter estimates. In Figure 3 it can be seen that the inter-individual differences in parameter estimates were smaller than the variability from measurement set to measurement set in the adaptive experiment. Parameter estimations of the two procedures and the different subjects are comparable (see Fig. 5). This is expected, because all the subjects were young, normal-hearing participants and should therefore not differ substantially in their thresholds. Future studies with hearing-impaired subjects are expected to reveal the full potential of the MoBES procedure by providing individual differences in the model parameter estimations.

Besides the focus of more efficient diagnostic measurements, one key advantage of using the MoBES procedure is the way it provides the researcher with a deeper understanding of the model in use. When comparing the selected stimulus conditions (see the bottom panel in Fig. 4) to the changes in the model prediction in Figure 1, it becomes obvious which stimulus conditions provide the most information about each of the model parameters. $N_{\pi}S_{\pi}$ is chosen, as it only depends on σ_{mon} . The frequently chosen condition

$N_{\rho=0.75}S_{\pi}$, for example, mainly informs about σ_{bin} . However, the more complex the models are, the more difficult it is to comprehend these relationships. Even when not using the MoBES module to steer the measurement, using it in the piloting phase of an experiment might add valuable knowledge about the inner mechanics of the model or which conditions should be measured in the main part of the experiment.

To finish, we note that the model-based steering procedure presented in this study is a useful tool for future research on auditory diagnostics. Characterization of individuals in terms of abstract parameters that influence hearing-aid fitting or maybe the choice of a hearing support device type is possible – at least in theory. Scientifically, both the likelihood-based fitting and the model-based steering foster a deeper understanding of the models in use. The procedure also offers insights into its interaction with fitting tools, measurement procedures, and subject peculiarities that are not captured by the model. Specifically, as argued by Herrmann and Dietz [11], tracing why the model chooses certain stimuli and in which order, is highly informative, even for an improvement of conventional manual measurement selection. It also facilitates a deeper understanding of the impact of each model parameter in general, and of each parameter's discretization steps. The procedure thus provides new perspectives for the design of diagnostic models and experiments.

5 Conclusion

The aim of this study was to test the feasibility of model-based experiment steering for the prediction of model parameters on the example of a tone-in-noise detection experiment. We showed that the procedure can be used to estimate model parameters more time-efficiently than a standard adaptive method. Thus, in the future, it can be used to assist in linking data to the underlying pathology, or to a parametric description of the individuals' abilities. The distant goal of diagnosing the causes of a person's hearing impairment has not yet been achieved because auditory models have either too many parameters or miss out on some diagnostically relevant aspects. However, the procedure already enables a deeper understanding of the model used and the impact of each model parameter. This is particularly important when working with more complex models. Furthermore, the procedure is not limited to audiological diagnostics, but can also be used in various fields other than audiology.

Conflict of interest

The authors declared no conflicts of interests.

Data availability statement

The Matlab code and data analyzed in this study are available online [21].

Supplementary figures

Supplementary figures are available at <https://acta-acustica.edpsciences.org/10.1051/aacus/2023064/olm>.

Figure S1: Compound likelihood (mean over the other three parameters and all stimulus parameters after setting the maximum of each trial to zero) for each of the four model parameters in the upper four panels for subject S1. The stimuli chosen by the procedure across trials are shown in the bottom panel. The dashed black line indicates the end of the initialization blocks and the start of the model steering with trial number 113.

Figure S2: Compound likelihood (mean over the other three parameters and all stimulus parameters after setting the maximum of each trial to zero) for each of the four model parameters in the upper four panels for subject S2. The stimuli chosen by the procedure across trials are shown in the bottom panel. The dashed black line indicates the end of the initialization blocks and the start of the model steering with trial number 113.

Figure S3: Compound likelihood (mean over the other three parameters and all stimulus parameters after setting the maximum of each trial to zero) for each of the four model parameters in the upper four panels for subject S3. The stimuli chosen by the procedure across trials are shown in the bottom panel. The dashed black line indicates the end of the initialization blocks and the start of the model steering with trial number 113.

Figure S4: Compound likelihood (mean over the other three parameters and all stimulus parameters after setting the maximum of each trial to zero) for each of the four model parameters in the upper four panels for subject S5. The stimuli chosen by the procedure across trials are shown in the bottom panel. The dashed black line indicates the end of the initialization blocks and the start of the model steering with trial number 113.

Acknowledgments

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme grant agreement no. 716800 (ERC Starting Grant to Mathias Dietz).

References

1. S. Hoth, I. Baljic: Current audiological diagnostics. GMS Current Topics in Otorhinolaryngology, Head and Neck Surgery 16 (2017) Doc09.
2. M.R. Panda, W. Lecluyse, C.M. Tan, T. Jürgens, R. Meddis: Hearing dummies: individualized computer models of hearing impairment. International Journal of Audiology 53, 10 (2014) 699–709.
3. R. Meddis: Auditory-nerve first-spike latency and auditory absolute threshold: a computer model. Journal of the Acoustical Society of America 119, 1 (2006) 406–417.
4. B. Sackmann, E. Dalhoff, M. Lauxmann: Model-based hearing diagnostics based on wideband tympanometry measurements utilizing fuzzy arithmetic. Hearing Research 378 (2019) 126–138.

5. S. Verhulst, A. Altoè, V. Vasilkov: Computational modeling of the human auditory periphery: auditory-nerve responses, evoked potentials and hearing loss. *Hearing Research* 360 (2018) 55–75.
6. J. Klug, L. Schmors, G. Ashida, M. Dietz: Neural rate difference model can account for lateralization of high-frequency stimuli. *Journal of the Acoustical Society of America* 148, 2 (2020) 678.
7. R. Plomp: Auditory handicap of hearing impairment and the limited benefit of hearing aids. *Journal of the Acoustical Society of America* 63, 2 (1978) 533–549.
8. M. Buhl, A. Warzybok, M.R. Schädler, T. Lenarz, O. Majdani, B. Kollmeier: Common Audiological Functional Parameters (CAFPAs): statistical and compact representation of rehabilitative audiological classification based on expert knowledge. *International journal of audiology* 58, 4 (2019) 231–245.
9. S.K. Saak, A. Hildebrandt, B. Kollmeier, M. Buhl: Predicting common audiological functional parameters (CAFPAs) as interpretable intermediate representation in a clinical decision-support system for audiology. *Frontiers in Digital Health* 2 (2020). 596433.
10. T. Brand, B. Kollmeier: Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *Journal of the Acoustical Society of America* 111, 6 (2002) 2801–2810.
11. S. Herrmann, M. Dietz: Model-based selection of most informative diagnostic tests and test parameters. *Acta Acustica* 5 (2021) 51.
12. R. Sanchez Lopez, F. Bianchi, M. Fereczkowski, S. Santurette, T. Dau: Data-driven approach for auditory profiling and characterization of individual hearing loss. *Trends in Hearing* 22 (2018) 2331216518807400.
13. J. Encke, M. Dietz: A hemispheric two-channel code accounts for binaural unmasking in humans. *Communications Biology* 5, 1 (2022) 1122.
14. I.J. Hirsh: The influence of interaural phase on interaural summation and inhibition. *Journal of the Acoustical Society of America* 20, 4 (1948) 536–544.
15. D.E. Robinson, L.A. Jeffress: Effect of varying the interaural noise correlation on the detectability of tonal signals. *Journal of the Acoustical Society of America* 35, 12 (1963) 1947–1952.
16. H. Levitt: Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America* 49, 2 (1971) 467–477.
17. K. Taylor, D. Rohrer: The effects of interleaved practice. *Applied Cognitive Psychology* 24, 6 (2010) 837–848.
18. X.D. Song, B.M. Wallace, J.R. Gardner, N.M. Ledbetter, K. Q. Weinberger, D.L. Barbour: Fast, continuous audiogram estimation using machine learning. *Ear and Hearing* 36, 6 (2015) e326–e335.
19. J. Encke, M. Dietz: Statistics of the instantaneous interaural parameters for dichotic tones in diotic noise (N0S ψ). *Frontiers in Neuroscience* 16 (2022) 1022308.
20. M.R. Leek: Adaptive procedures in psychophysical research. *Perception & Psychophysics* 63 (2001) 1279–1292.
21. A. Dietze: Matlab code and datasets for: Auditory model-based parameter estimation and selection of the most informative experimental conditions. 2023. <https://doi.org/10.5281/zenodo.7863204>.

Cite this article as: Dietze A. Reinsch A-L. Encke J. & Dietz M. 2024. Auditory model-based parameter estimation and selection of the most informative experimental conditions. *Acta Acustica*, 8, 3.